# Symposium Connects Government Problems with State-of-the-Art Network Science Research

*By Rajmonda S. Caceres and Benjamin A. Miller*

Network science has grown significantly in the last several years as a field at the intersection of mathematics, computer science, social science, and engineering. Topics of interest include modeling and analysis of network phenomena, large-scale computation and data management, models for information and epidemic spreading through networks, and inference of information about entities based on observable connections. While basic research is focused on developing understanding in each of these areas, in a practical setting the ultimate goal is to exploit this understanding to achieve some application-specific objective.

As a Federally Funded Research and Development Center for the United States Department of Defense, MIT Lincoln Laboratory has a number of research efforts driven by graph exploitation tasks. In diverse mission areas such as cyber security, counterterrorism/counterinsurgency, air traffic control, and bioengineering, the data of interest are inherently interconnected, and using a network or graph representation for the data enables an additional level of insight not available when considering the data independently.

In 2010, as part of a research effort funded by the Office of Naval Research, Lincoln Laboratory staff set out to build a community of interest through a symposium specifically focused on exploitation of graph data. The goal was to bring together academic researchers, industry practitioners, and end users to discuss problems of interest to the US Government, and match these with the state-of-the-art models and techniques developed in the network science research community. Since its inception, the Graph Exploitation Symposium (GraphEx) has been held annually as a meeting to facilitate this interaction.

GraphEx 2015—held 16–17 July in Dedham, Massachusetts—was organized by a committee comprised of several Lincoln Laboratory staff and recent symposium participants. Oral and poster presentations were all by invitation, and included speakers from academia, government, national laboratories, and industry. The participants included network science researchers and practitioners with expertise in machine learning, statistics, security, signal processing, and big data analysis. Topics covered over the two-day meeting ranged from exploitable theoretical insights to specific, mission-focused problems with a graph-theoretic flavor, with the following themes outlining the current research frontier within the network science field.

*Controllability of large-scale complex networks:* Many human engineered systems—including computational infrastructures, transportation systems, and electrical power grids—behave like large-scale complex networks with different layers and components interacting in non-trivial ways. The susceptibility of such systems to local shocks is amplified via network cascading effects. Currently, we lack the ability to quantify vulnerability and resilience of large-scale complex networks at the system level. Symposium speakers put forward several suggestions to address this important challenge:

- Combine network science approaches with existing rich methodologies from the fields of control theory and decision theory to model dynamically changing demands on (and capacity of) the engineered complex system.
- Rigorously measure, model and assess the system architecture at different levels and under different constraints and disruptions.
- Design complex systems that are dynamic and stable by adaptively allocating resources/services through closed-loop feedback channels (e.g. software defined wireless networks, intelligent transportation systems).
- Design complex network control mechanisms that can robustly handle stochastic cascades on the network and be able to localize control effects.

*User-centric algorithms:* Combining user-centric, private, rich data with global, incomplete, publically available data was the theme of several talks, from estimating systemic risk in transportation networks to designing recommender systems on social networks. Mathematical models and algorithms that can handle a seamless and efficient integration of the two data regimes have implications beyond accuracy improvements on inference tasks. Such models can become enabling technologies as we enter the new era of data democratization and user self-awareness and empowerment.

*Noise and interference in networks:* Exploiting graphs in the presence of noise and interference is another recent topic of interest in the community, and several presentations touched on this point. In practice, users typically have some prior knowledge (or domain expertise) suggesting probable structure within the network. Algorithms should be developed in consideration of the fact that some of the structure is uninteresting. Material discussed during the symposium included experimental design in the presence of interference in order to optimize inference ability from the measurements. As community detection relies on metrics that are sensitive to noise in the data, metrics that are resilient to noisy observations was another topic of interest.

*Network analysis in adversarial settings:* A variant graph analysis with noise and interference is the case where the data are specifically manipulated by an adversary to counter the exploitation task. This area has significant implications for tasks in cyber security and counterterrorism, where a common objective is to uncover a subgraph of interest in which actors are deliberately covert. Material presented on this topic included models for attacker behavior and classes of methods to mitigate the effects of purposeful data corruption.

*Multi-modal, multi-layer networks:* As many graphs come from multiple sources, fusion of information across multiple observations was another theme that ran through several presentations. Analyzing dynamic graphs and multigraphs has become a necessity, as networks are more frequently observed as objects that change over time and have different connections when viewed through different media. Challenges in this area include developing models for evolution of graphs over time, optimal fusion of information across modalities, and quantification of the benefits and limitations of inference across observations.

*Big data analysis and management:* Graphs are frequently extracted from extremely large datasets, and dealing with data of this scale and variety is a challenge inherent in modern network analysis. Several presentations touched on these issues, including personalized recommendations from global data and exploitation of open source and social media data for disaster response. From the big data management perspective, there are currently active efforts toward implementation of key computational kernels for graph exploitation algorithms natively in a large-scale database system.

Overall, the 6th GraphEx Symposium continued its successful tradition of bringing together unique perspectives from research, applications, and operations, and influencing directions of basic network science research in support of current, critical technological needs. As capabilities and technologies evolve, it is the intention to maintain GraphEx as a venue for ensuring this research-to-practice connectivity.

*Information about the Lincoln Laboratory's Graph Exploitation Symposium, including proceedings of past meetings, can be found at https://events.ll.mit.edu/graphex/. Rajmonda S. Caceres and Benjamin A. Miller are Technical Staff in the Cyber Analytics and Decision Systems Group at MIT Lincoln Laboratory.* *This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.*